International Journal of Information and Computer Science
IJICS Volume 1, Issue 2, May 2012 PP. 21-29
ISSN(online) 2161-5381
ISSN(print) 2161-6450
www.iji-cs.org

Binocular Stereo Localization Using Two Pan-Tilt-Zoom Cameras

Lining Gao[1], Qiuqi Ruan[2]
Institute of Information Science, Beijing Jiaotong University, Beijing, China
Email:10120383@bjtu.edu.cn

(Abstract) The traditional binocular stereo vision is mainly based on stationary cameras. Compared with stereo vision using dual-stationary cameras, binocular stereo vision using two pan-tilt-zoom cameras is more challenging since the camera parameters may change when the camera motion. Few literatures focused on this work. In this paper, we propose an improved pan-tilt-zoom camera model, which parameterizes camera projection with the camera's pan-tilt angles. Two different methods are presented to calibrate the camera model. Stereo vision using two pan-tilt-zoom cameras can be converted to traditional stereo vision via a homography matrix defined in the proposed model. Therefore, a lot of algorithms used in traditional converged stereo vision can be valid for stereo vision using pan-tilt-zoom cameras. We use the binocular stereo to get three-dimensional information of object on the dual-arm mobile robot platform. The experiment results show that the proposed approach is promising and is able to locate the targets with high accuracy.

Keywords: Camera Calibration; Stereo Localization; Dual Pan-tilt-zoom Cameras; Dual-arm Mobile Robot

1. INTRODUCTION

In recent years, there have been considerable interest in the research of intelligent mobile robot. Vision systems are used to get three-dimensional information of object, which are critical to autonomous operation and navigation of robot. Most vision-based approaches to extracting object information can be classified into two categories. The first one is using monocular camera with known scene information, which is very difficult if no knowledge of reference scale is provided; and the second one is using binocular camera, which always has two cameras as one equipment for convenience in stereo rectification and matching. The second one is similar to human eyes system, compared with monocular system, it tends to be more stable and better behaved. In addition, one does not need to worry about the scale ambiguity presented in monocular camera case. Our study belongs to the binocular vision.

Binocular stereo vision is one of the most significant embranchments of computer vision [1], and extraction of three-dimensional information of the scene from stereo images is a challenging issue that has been studied by the computer vision community for decades [2]. Traditional stereo vision research usually use stationary cameras, which is relatively simpler because the internal and external parameters of the camera are constant after the stereo system are constructed. Since pan-tilt-zoom (PTZ, for short) cameras' pose can be controlled by pan, tilt and zoom parameters, it allows us to obtain multi-view-angle and multi-resolution scene information using fewer cameras. However, stereo vision using dual-PTZ-camera (see **Figure 1**) is much more challenging as the internal and external parameters of camera can be changed in utility. The process of calibration is to determine the internal and external parameters of a camera from a number of correspondences between 3D points and their projections onto one or multiple images. Generally, this process is accomplished using a calibration plane with a checkerboard or other known marker pattern. Previous works on active camera calibration have mostly been done in a laboratory setup using calibration targets and LEDs or at least in a controlled environment. Some of these include active zoom lens calibration by Willson et. al. [3] // [4], self-calibration from purely rotating cameras by deAgapito [5], and more recently pan-tilt camera calibration by Davis et.Al [6]. The common ground of these approaches is that they belong to single camera calibration and do not make use of the pan-tilt angles information, which can be obtained from the camera. Recently, Dingrui Wan [7] proposed a novel camera calibration and stereo rectification method for dual-PTZ-camera system, which is essential to increase the efficiency of stereo matching greatly. Kumar et al. [8] presented which can locate a moving target in a complex environment based on two PTZ cameras. A

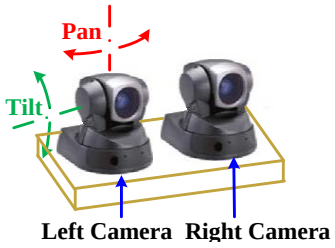


Figure 1. Dual-PTZ-camera system

look-up-table (LUT, for short) of rectification matrices is constructed off line for computing the rectification transformations for arbitrary camera positions in real time. Our work is similar to Wan and Zhou [7] and Kumar [8] in that they are both based on dual-PTZ-camera system and make use of the pan-tilt angles information obtained from the PTZ camera.

In this paper, a novel approach of target localization based on two PTZ cameras is presented. First, we discuss the PTZ camera model with the rotated camera and propose a new PTZ camera model by adding the camera's pan-tilt angles as parameters of the camera projection matrix. Second, two different methods based on existing camera calibration methods are provided to calibrate the proposed camera model. Third, stereo localization using two pan-tilt-zoom cameras can be converted to traditional stereo Localization using two aligned cameras by a homography matrix defined in the proposed model. Finally, a software system developed by VC++ on the dual-arm mobile robot platform is presented to show the robot tracking and grabbing experiment results. Compared with [7][8] , our dual-PTZ-camera stereo is better for that many algorithms for traditional stereo vision can be used. The stereo localization accuracy is good enough for target operating in our mobile robot.

The paper is organized as follows: Section 2 describes the traditional and improved camera model respectively, in section 3 both simple and complicated methods for calibrating the model are presented. In section 4, we describe the stereo localization using the dual-PTZ-camera. Experimental results are provided in section 5. We conclude with discussions in Section 6.

CAMERA MODEL

1.1. Traditional Camera Model

In the traditional pin-hole model (see **Figure 2**) for the perspective camera, a point X in 3D projective space P^3 projects to a point x on the 2D projective plane P^2 (the image plane). This can be represented by a mapping $f : P^3 \to P^2$ such that $x = PX$ (Note: the equality relation when applied to homogeneous vectors really means equality up to a non-zero scale factor), where $x = [x, y, 1]^T$ is the image point in homogeneous coordinate, $X = [X, Y, Z, 1]^T$ is the world point, and P is the 3×4 camera projection matrix. The matrix P is a rank-3 matrix which may be decomposed as $P = K[R \mid -Rt]$, where the rotation R and the translation t are known as external parameters which represent the Euclidean transformation between the camera and the world coordinate systems, and K is nonsingular 3×3 upper triangular matrix including five parameters, which encodes the internal parameters of the camera in the form:

$$K = \begin{bmatrix} \alpha f & s & p_u \\ 0 & f & p_v \\ 0 & 0 & 1 \end{bmatrix} \quad (1)$$

where f is the focal length and α is the aspect ratio. The principal point is (p_u, p_v) and s is a skew parameter which is a function of the angle between the horizontal and vertical axes of the sensor array. The principal point (p_u, p_v) and focal length f depend only on the camera zoom z, we usually assume the skew parameter $s = 0$, hence the internal parameters K is

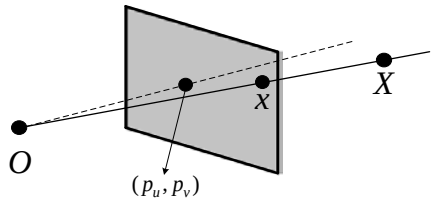


Figure 2. The pin-hole camera model

constant for a particular zoom z, K may be written as:

$$K(z) = \begin{bmatrix} \alpha f(z) & s & p_u(z) \\ 0 & f(z) & p_v(z) \\ 0 & 0 & 1 \end{bmatrix} \quad (2)$$

We usually solve the following optimization problem to calculate $K(z)$:

$$\arg\min \sum_{i=1}^{M} \left\| x_i - \hat{x}_i \right\|^2 \quad (3)$$

where $\hat{x}_i = K(z)[R \mid -Rt]X^i$, X^i represents a 3D point, x^i represents corresponding 2D projection point, which is detected by feature-based approaches. Commonly used are the Harris corners [9] or the more stable SIFT features [10]. **Eq.3** means to minimize the Euclidean distance between corresponding points, the general method used is Levenberg-Marquardt least-squares parameter estimation method [11]. In our method, the zoom z is fixed, we do not focus on this problem, more details about zoom parameters estimation can be found in [7][12].

1.2. Improved Camera Model

Now we will consider what will happen to the camera model when the camera is rotated either by pan or tilt motion. As indicated above, the internal parameters depends only on the zoom parameter, hence it is constant when rotating the camera. Different to internal parameters, the external parameters will change with the camera motion. It means that the solved parameters under one camera pan-tilt parameter setting are not valid under another. Therefore, we try to find some constant parameters and model the variable parameters with pan-tilt angles.

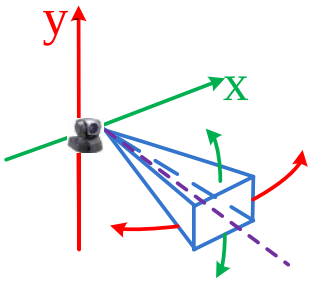
For convenience, the origin of the world coordinate system chosen to be the location of the camera, *i.e.* the translation vector $t = 0$, and the camera pan-tilt motion is considered as pure rotation without translation (see **Figure 3**). Since the last column of P is always 0, the projection matrix can be written as $P = KR$. Let x and x' be the images of X taken by a rotating camera under two different pan-tilt parameter settings. Then x and x' are related to X as $x = PX = KRX$ and $x' = P'X = K'R'X$, consider that K is constant when rotating the camera, then $x' = KR'R^{-1}K^{-1}x$ .Let $R_{rot} = R'R^{-1}$ represents the relative camera rotation about its projection

center between the two views, then the equation reduces to:

$$x' = KR_{rot}K^{-1}x \quad (4)$$

Let $H_{rot} = KR_{rot}K^{-1}$ represent the homography, **Eq.4** can be reduced to

$$x' = H_{rot}x \quad (5)$$



$$x = P'X = H_{rot}(\psi,\theta)P_0X$$

Figure 3. The pan-tilt motion of PTZ camera

Similar to [13], we conclude the following result from **Eq.5**. Given a pair of images taken by cameras with the same internal parameters from the same location, then there exist 2D projective transform, represented by matrices H_{rot}, taking one image to the other.

Consider $x' = P'X$ and $x = PX$, from **Eq.5** we can give the following result :

$$P' = H_{rot}P \quad (6)$$

Where P' and P are camera projection matrix under different pan-tilt parameters setting and H_{rot} is same to the one defined in **Eq.5.**

Combine the traditional camera model $x = PX$ with **Eq.6**, we propose the improved PTZ camera model as follows:

$$x = H_{rot}(\psi,\theta)P_0X = KR_{rot}(\psi,\theta)K^{-1}P_0X \quad (7)$$

Where $X = [X,Y,Z,1]^T$ is a point in 3D projective space, and $x = [x,y,1]^T$ is the projection point of X on image in homogeneous coordinate, P_0 is the projection matrix when the camera's parameters setting is $(Pan = 0, Tilt = 0)$, $R_{rot}(\psi,\theta)$ represents the relative camera rotation between the view when camera's parameter setting is $(Pan = 0, Tilt = 0)$ and the view when camera's parameter setting is $(Pan = \psi, Tilt = \theta)$,and $H_{rot}(\psi,\theta) = KR_{rot}(\psi,\theta)K^{-1}$ is the homography matrix, which can take one image captured when the camera's parameters setting is $(Pan = \psi, Tilt = \theta)$ to the image captured when the camera's parameter setting is $(Pan = 0, Tilt = 0)$.

CALIBRATING THE MODEL

The improved camera model consists of the projection matrix P_0 and the homography matrix H_{rot}. First, we use traditional camera calibration method to compute P_0. Second, two methods are proposed to compute the homography matrix H_{rot}, which depends on the pan-tilt parameters.

1.3. Compute the projection matrix

Due to its importance, much work has been done in the field of camera calibration with all kinds of proposed approaches. The common practice for camera calibration is to collect a set of correspondences between 3D points and their projections on image plane [15]. Camera parameters can be determined by solving:

$$\arg\min_{\varphi}\{P(\varphi, X_{3D}) - X_{2D}\} \quad (8)$$

Where P is the camera model that defines the projection of points onto the image plane, φ is the set of camera parameters to be determined, X_{3D} is the vector of 3D feature locations, and X_{2D} is the vector of corresponding image plane observations. Since φ often includes radial lens distortion terms in addition to extrinsic geometric pose, minimizing this equation is usually framed as a non-linear search problem.

The two most common techniques for camera calibration are those of Tsai [16] and Zhang [17].While the method of Tsai has the advantage that it can handle both coplanar and non-coplanar input point. The easiest and most practical approach is to use a calibration grid or checkerboard of coplanar points. When the input points lie on a single plane, it is wise to have multiple input images containing different planar grid orientations in order to ensure a robust calibration. Zhang's calibration method strictly enforces these conditions, requiring multiple images of a planar calibration grid. Compared with classical techniques which use expensive equipment such as two or three orthogonal planes, Zhang's technique is easier to use and more flexible.

Feature detection is an important step for camera calibration. Traditional algorithm for detecting X-corners first finds their pixel positions by Harris detector based on a hessian matrix looking for the auto-correlation matrix:

$$M = \begin{pmatrix} \left(\frac{\partial I}{\partial x}\right)^2 \otimes w & \left(\frac{\partial I}{\partial x}\cdot\frac{\partial I}{\partial y}\right)\otimes w \\ \left(\frac{\partial I}{\partial x}\cdot\frac{\partial I}{\partial y}\right)\otimes w & \left(\frac{\partial I}{\partial y}\right)^2 \otimes w \end{pmatrix} \quad (9)$$

Where w is a Gauss smoothing operator. Harris corner detector is expressed as:

$$R = \det(M) - \lambda(trace(M))^2 \quad (10)$$

The X-corner is just the local peak point of R. Chen [18] proposed a new sub-pixel detector for X-corners, which is much simpler than the traditional sub-pixel detection algorithm.

In this paper, we use Chen's method to detect sub- pixel corners and Zhang's method to calculate the projection matrix P_0 when the camera's parameters setting is $(Pan = 0, Tilt = 0)$.The main steps are as follows:

- Print a pattern and attach it to a planar surface.
- Take a few images of the model plane under different orientations by moving the plane.
- Detect feature points in the images using Chen's method.
- Estimate the five internal parameters and all the external parameters using the closed-form solution.
- Estimate the coefficients of the radial distortion by linear least-squares method.
- Refine all parameters by maximum likelihood estimation.

1.4. Compute the homography matrix

The homography matrix that depends on the pan-tilt parameter setting consists of rotation R_{rot} and matrix K. The internal parameters K had been solved while computing the projection matrix given in section 3.1. Therefore, the critical work is to solve the rotation R_{rot}.There are various methods to parameterize rotations. A rotation can be represented by a unit quaternion in the method of quaternions. This method of parametrization has the disadvantage of using four parameters per rotation and requiring the quaternions to be renormalized at each step. Better method to parametrize rotations is Eulerian angles, which has the advantage that a rotation is parametrized by the minimum of three parameters. **Figure 4** shows a rotation represented by Eulerian angles, where θ, ψ and φ are the rotation angles around the x-axis, y-axis and z-axis of the reference camera coordinate respectively. When the x-axis is the rotation axis, R_{rot} can be written as:

$$R_X = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{bmatrix} \tag{11}$$

In a similar manner, when the y-axis or z-axis is the rotation axis, R_{rot} can be written as:

$$R_Y = \begin{bmatrix} \cos\psi & 0 & -\sin\psi \\ 0 & 1 & 0 \\ \sin\psi & 0 & \cos\psi \end{bmatrix} \quad \text{or} \quad R_Z = \begin{bmatrix} \cos\varphi & \sin\varphi & 0 \\ -\sin\varphi & \cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{12}$$

We choose a simple PTZ camera model similar to [7][12]. Our model assumes that the pan and tilt axes of rotation are orthogonal, aligned with the image plane, and the axes of rotation intersect the optical center of the camera imaging system, *i.e.* the pan and tilt axes of rotation correspond to y-axis and x-axis of the reference camera coordinate respectively (see **Figure 4**). Based on the above-mentioned assumption, we present a simple method to solve the rotation matrix as follows:

$$R_{rot} = R_X \cdot R_Y = \begin{bmatrix} \cos\psi & \sin\psi\sin\theta & -\sin\psi\cos\theta \\ 0 & \cos\theta & \sin\theta \\ \sin\psi & -\cos\psi\sin\theta & \cos\psi\cos\theta \end{bmatrix} \tag{13}$$

where ψ and θ are the pan and tilt angles respectively, which can be obtained from the camera system conveniently.

For more general model that violate the assumption, a complicated method will be proposed later. Reference [14] mentioned that when a pan-tilt camera was assembled, it was difficult to ensure that the axes of rotation intersect the optical center of the camera imaging system, and the axes were unlikely to be exactly aligned. In order to model the actual motion of the camera, new parameters were introduced into the camera model in [14], the proposed model approximates the actual camera geometry more closely, it can be written as:

$$x = PT_{pan}R_{pan}T_{pan}^{-1}T_{tilt}R_{tilt}T_{tilt}^{-1}X \tag{14}$$

Where R_{tilt} is a 3x3 rotation matrix which rotates tilt angles around a vector in the direction of the tilt-axis. R_{pan} is analogous. These axes do not necessarily pass through the origin, and T_{pan} and T_{tilt} represent translation vectors from the origin to each axis. Thus the projection of a point in space onto the image plane can be found as a function of current camera pan and tilt parameters. T_{pan} and T_{tilt} are negligible in our camera system, hence we propose a simpler model here:

$$x = PR_{pan}R_{tilt}X \tag{15}$$

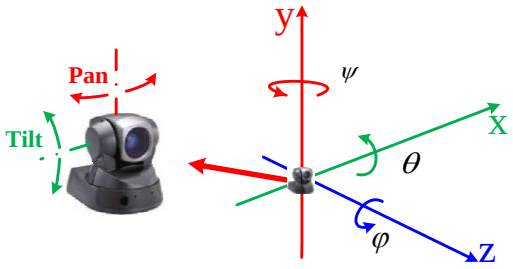


Figure 4. The camera's rotation represented by Eulerian angles

Compared with the simple model mentioned in **Eq.13**, the model in **Eq.14** does not assume that the pan and tilt axes of rotation correspond to y-axis and x-axis of the reference camera coordinate respectively, *i.e.* given the camera's parameters setting $(Pan = \psi, Tilt = \theta)$, we need to estimate the actual angles around the x-axis, y-axis and z-axis of the reference camera coordinate respectively. Consider that $R_{rot}(\psi,\theta) = R'R^{-1}$, where R and R' are the rotation matrix when the camera parameter setting are $(Pan = 0, Tilt = 0)$ and $(Pan = \psi, Tilt = \theta)$, We can solve the following optimization problem to calculate R and R':

$$\arg\min \sum_{i=1}^{M} \left\| x_i - \hat{x}_i \right\|^2 \tag{16}$$

Where $\hat{x}_i = KR(\psi,\theta,\varphi)X^i$, X^i represents a 3D point, x^i represents corresponding 2D projection point detected by feature-based approaches, ψ,θ,φ are the parameters to be determined, others are known. After solving R and R', $R_{rot}(\psi,\theta)$ can be calculated by matrix multiplication. We can calculate the corresponding $R_{rot}(\psi,\theta)$ for each parameter setting $(Pan = \psi, Tilt = \theta)$ off line, therefore a LUT of $R_{rot}(\psi,\theta)$ based on pan-tilt parameters can be constructed. The main steps to construct the LUT are as follows:

- Sample the pan-tilt angles for the left camera using the angle range as follows:
  $$Pan^L = P^L_{\min} : 1 : P^L_{\max}, Tilt^L = T^L_{\min} : 1 : T^L_{\max}$$
- In a similar manner, sample the pan-tilt angles for the right camera using the angle range as follows:
  $$Pan^R = P^R_{\min} : 1 : P^R_{\max}, Tilt^R = T^R_{\min} : 1 : T^R_{\max}$$
- Capture at least one image for each combination of (Pan^L, $Tilt^L$, Pan^R, $Tilt^R$) for the left and right camera. The total number may be $(P_{\max} - P_{\min} + 1) \times (T_{\max} - T_{\min} + 1)$ for each camera.
- For each combination of (Pan^L, $Tilt^L$, Pan^R, $Tilt^R$) satisfied that the two cameras share at least 30% of their

view, calculating the corresponding $R_{rot}(\psi,\theta)$ for each camera.

The rotation $R_{rot}(\psi,\theta)$ can be solved by either the simple method or the complicated LUT method, and then the homomograhy matrix $H_{rot}(\psi,\theta)$ can be calculated by matrix multiplication as follows:

$$H_{rot}(\psi,\theta) = KR_{rot}(\psi,\theta)K^{-1} \quad (17)$$

2. STEREO LOCALIZATION

A point P in 3D space can be reconstructed by binocular stereo vision system as shown in **Figure 5**. P_L and P_R are the projection points of P imaged by two cameras C_L and C_R with optical centers O_L and O_R respectively. This problem can be solved by intersecting the rays from the optical centers O_L and O_R to projection points P_L and P_R. Actually, nonverged stereo are more common(see **Figure 6**). In the nonverged geometry, both camera coordinates axes are aligned and the baseline is parallel to the camera x coordinate axis. It follows that, for the special case of nonverged geometry, a point in space projects to two locations on the same scan line in the left and right camera images. The resulting displacement of a projected point in one image with respect to the other is termed as disparity ($d = X_L - X_R$). Given the distance between O_L and O_R, called the baseline B, and the focal length f of the cameras, depth at a given point may be computed by similar triangles as follows:

$$Z = \frac{bf}{X_L - X_R} = \frac{bf}{d}$$

hence,

$$\frac{b}{Z} = \frac{(b + X_R) - X_L}{Z - f} \quad (18)$$

Since depth is inversely proportional to disparity, there is obviously a nonlinear relationship between these two terms. When disparity is near 0, small disparity differences make for large depth differences. When disparity is large, small disparity differences do not change the depth too much. The consequence is that stereo vision systems have high depth resolution only for objects relatively near the camera.

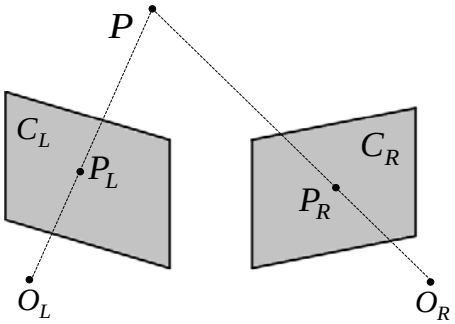


Figure 5. Point reconstruction in 3D space

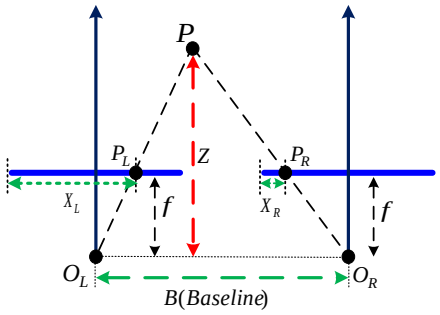


Figure 6. Nonverged stereo vision

It is easiest to compute the stereo disparity when the two image planes align exactly. Unfortunately, it is difficult to build stereo systems with nonverged geometry, since the two cameras almost never have exactly coplanar, row-aligned imaging planes. The goal of stereo rectification is to solve the above-mentioned problem, more details can be found in [15][19][20].

When both of the two cameras' parameter setting are $(Pan = 0, Tilt = 0)$, the traditional reconstruction method mentioned above can be used directly. More general cases, PTZ cameras' parameter setting do not satisfy the condition. In that situation, the homography matrix can be used to take the image captured when the cameras' parameter setting is $(Pan = \psi, Tilt = \theta)$ to the image captured when the cameras' parameter setting is $(Pan = 0, Tilt = 0)$, which is shown as **Eq.5**. After that, the traditional reconstruction method can be used.

3. EXPERIMENT RESULTS

All experiments were accomplished based on the ASR robot platform(see **Figure 7**). The robot consists of computer with a Windows XP operating system and VC++ development tools, vision system, motion system and dual-arm system. Our binocular stereo vision system consists of two SONY EVI-D31cameras, which are installed on the head of the robot. Both of the left and right camera can rotate in horizontal and vertical plane, angle range of pan is [-100,100], angle range of tilt is [-25,25]. The camera's maximal frame rate is 30. The image resolution is 640×480. Main steps for calibrating the camera are as follows:

- Print a pattern and attach it to a planar surface, the model plane contains a pattern of 6 × 9, and the size of the pattern is 25mm × 25mm.
- Adjust the parameters setting of the camera to $(Pan = 0, Tilt = 0)$.
- Take a few (actually are 18) images of the model plane under different orientations by moving the plane.
- Calibrate the internal parameters of both the left and right cameras using Zhang based method.
- Keep the plane fixed, construct the LUT of the homography matrix.

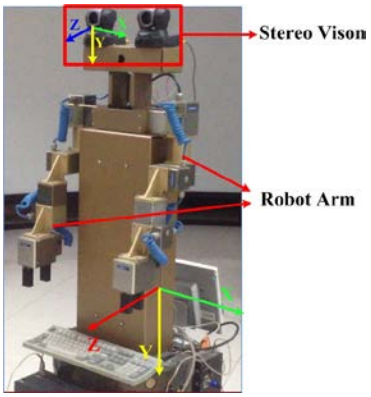


Figure 7. The dual-arm mobile robot

camera are close to the right one, but not exactly the same. The possible reasons may include both the calibration error and differences in manufacturing. We choose a world coordinate system paralleled to the camera plane, so the rotation matrix is almost an identity matrix.

Table 1. The results of internal camera parameterts.

parameters	α	f	p_u	p_v
Left camera	1.00004	549.127	333.398	236.993
Right camera	0.9995	546.247	333.258	213.622

The quality of camera calibration can be measured in terms of image plane projection error. This is essentially a measure of the camera model's ability to explain data. If the model is of high quality, the projection of target locations onto the image plane will fall closely to actual observations. **Figure 10** gives the projection error when using the improved

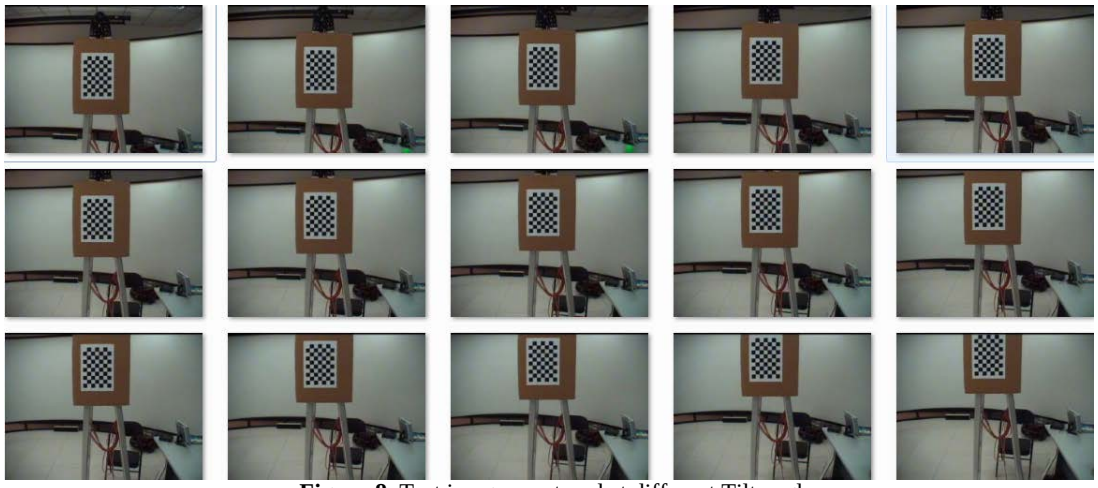


Figure 8. Test images captured at different Tilt angles

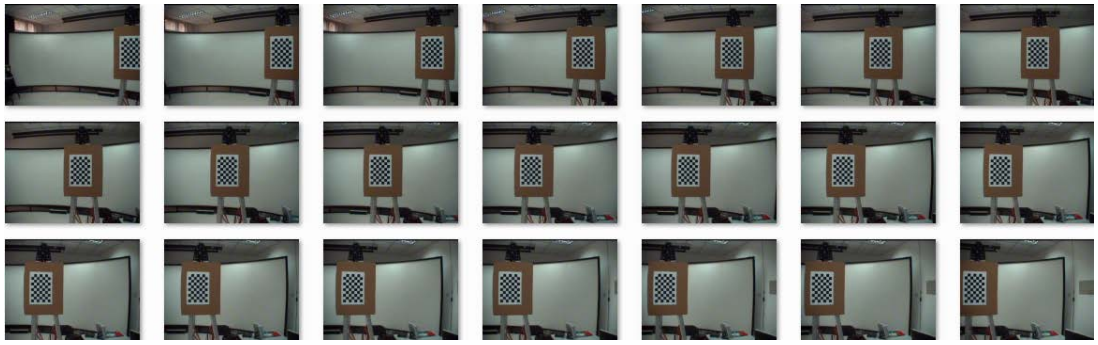


Figure 9. Test images captured at different Pan angles

In order to evaluate the effectiveness of our calibration methods, we obtained both calibration and test data. The test data was captured in an identical manner as the calibration data, while the two sets were kept separately. The test images were captured at different Tilt angles(See **Figure 8**) or different Pan angles(See **Figure 9**).

Results of the camera's internal parameters are shown in **Table 1**, and the external parameters when the camera's parameters setting is $(Pan = 0, Tilt = 0)$ are shown in **Table 2**. From the table, we can see the internal parameters of the left

Table 2. The results of external camera parameterts.

parameters	R	T
Left camera	$\begin{bmatrix} 0.996 & -0.050 & 0.069 \\ -0.051 & -0.999 & 0.002 \\ 0.069 & -0.006 & -0.997 \end{bmatrix}$	[-41.38,73.46,563.32]
Right camera	$\begin{bmatrix} 0.993 & -0.029 & 0.113 \\ -0.030 & -0.999 & 0.006 \\ 0.112 & -0.009 & -0.994 \end{bmatrix}$	[-226.02,80.28,573.51]

model while the homomograhy matrix is computed by thesimple method, the error level is large when the pan-tilt angle is far from 0, hence it is effective when the rotation angle is small. **Figure 11** gives the result when using the improved model while the homomograhy matrix is computed by the LUT method**,** it has a high quality because much work has been done manually when constructing the LUT . **Figure 12** gives the result using traditional camera model, we only use the calibration data when the camera's parameters setting is $(Pan = 0, Tilt = 0)$, the error level is unacceptable because the external camera parameters will change when the camera motion. The comparison of average projection error for different pan or tilt angles under three conditions are shown in **Table 3**, where notation 1 represents the condition using the traditional model, notation 2 represents the condition using the improved model while homomograhy matrix is computed by the simple method, notation 3 represents the condition using the improved model while homomograhy matrix is computed by the LUT method.

Table 3. The average projection error under three conditions.

	1	2	3
Error /pixel (Pan)	123.37	5.52242	0.209294
Error /pixel (Tilt)	74.527	3.32231	0.203621

The external parameters represent the relative position between two cameras. They were calculated when the camera parameters setting is $(Pan = 0, Tilt = 0)$. The translation vector is $T = [-211.04, -1.98, -3.64]$ and the rotation vector is $R = [-0.030, 0.007, 0.020]$, it means that the two cameras' baseline width is about 21cm and the rotation is almost zero. Results of stereo localization using the binocular stereo vision are given in **Figure 13**. The reconstructed points are close to the observed points. We define the absolute error as follows:

$$E_W = \frac{\sum_{i=1}^{N} \sqrt{(X_{ti} - X_i)^2 + (Y_{ti} - Y_i)^2 + (Z_{ti} - Z_i)^2}}{N} \quad (19)$$

(X_{ti}, Y_{ti}, Z_{ti}) is the reconstructed data. The average error is 0.728651cm.

Finally, we developed a software system by VC++ (See **Figure 14**).The system consists of 5 parts, image capture module, camera motion module, image processing module, robot motion module and robot arm module. In our software platform, the robot can interaction with human and obtain three-dimensional information of the target by stereo vision. Through the user interface, we can select the color of the target to grab, and three-dimensional information of the target can be computed in real time.

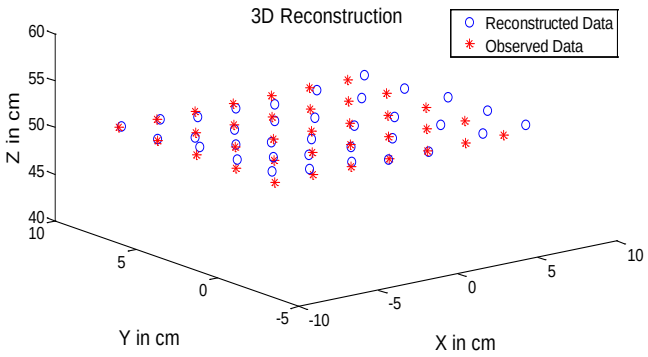


Figure 13. Observed data and reconstructed data in 3D coordinate

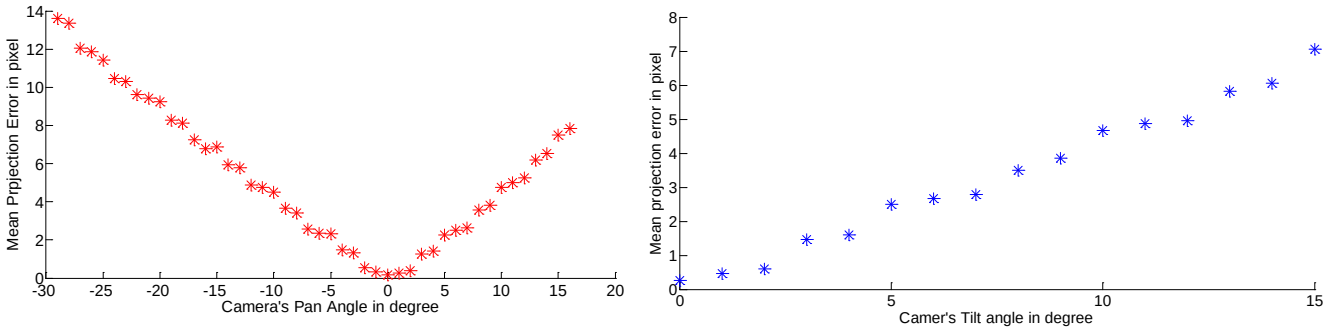


Figure 10. Mean projection error under the simple method condition and the test images are captured at different Pan or Tilt angles

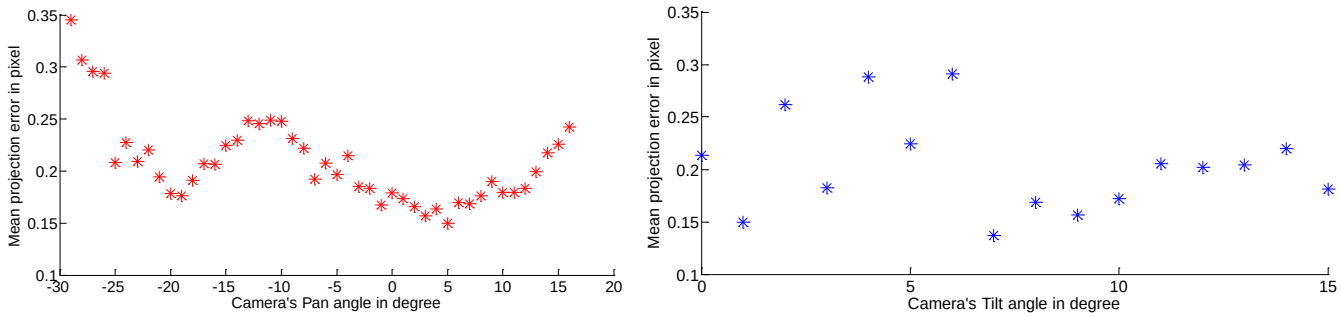


Figure 11. Mean projection error under the LUT method condition and the test images are captured at different Pan or Tilt angles

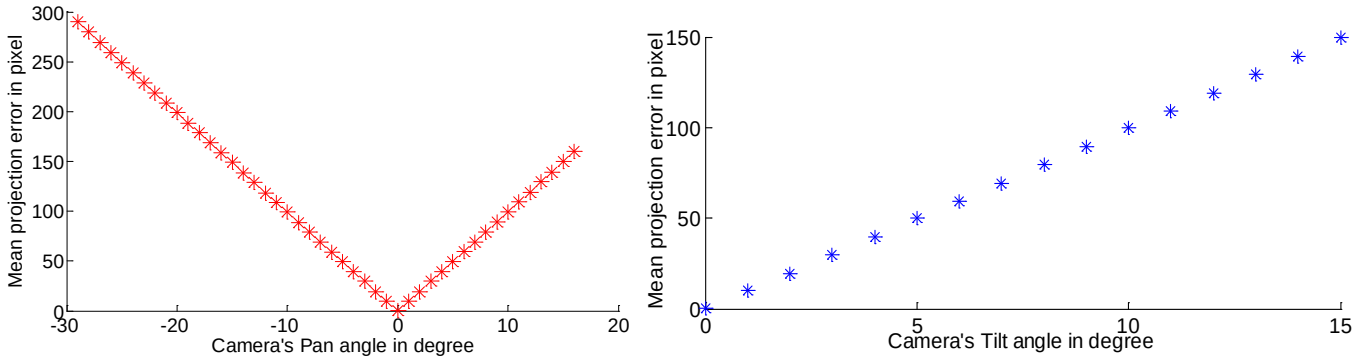


Figure 12. Mean projection error under the traditional model condition and the test images are captured at different Pan or Tilt angles

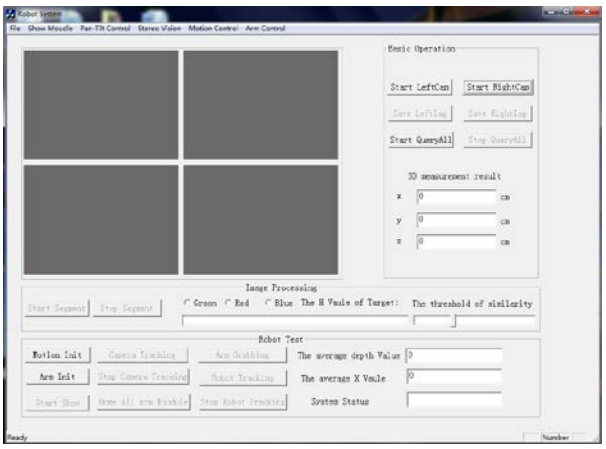


Figure 14. The software system

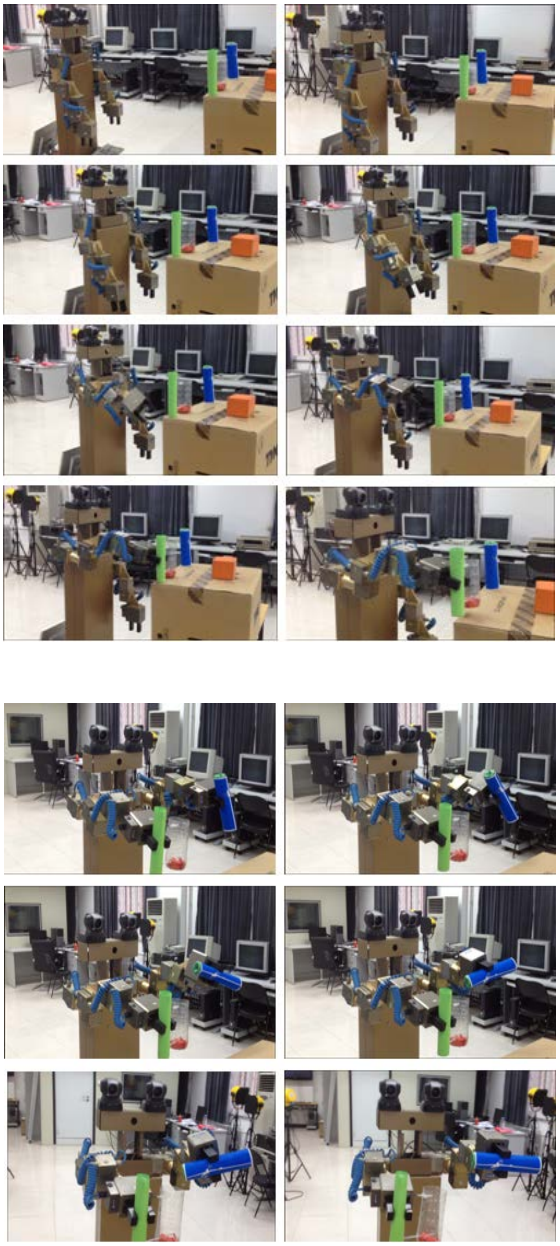


Figure 15. Intelligent tracking and grabbing of target

Intelligent tracking and grabbing of target are well performed on our robot platform. The results are shown in **Figure 15**, where green cylindrical is target 1 for right arm grabbing, and blue cylindrical is the target 2 for left arm grabbing.

4. CONCLUSIONS

An approach for target localization based on two PTZ cameras is presented in this paper. We proposed an improved camera model for PTZ camera. Two methods are presented to compute the homography matrix. Stereo vision based on two pan-tilt-zoom cameras can be converted to traditional stereo vision by the homography matrix defined in the proposed model. Therefore, traditional stereo algorithms can be used in dual-PTZ-camera system, which can enhance the flexibility of stereo vision system greatly. Experiment results show that the improved model provided in this paper is effective. The presented dual-PTZ-camera system can be applied to many fields, such as video surveillance and robot navigation.

5. ACKNOWLEDGEMENTS

This work is supported by Chinese National Natural Science Foundation (60973060) and Research Fund for Doctoral Program (200800040008).

REFERENCES

[1] D. Forsyth, J. Ponce, Computer Vision: A Modern Approach, Prentice Hall, 2003.

[2] M.Z. Brown, D. Burschka, G.D. Hager, Advances in computational stereo, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (8) (2003) 993–1008.

[3] R.G. Willson, Modeling and Calibration of Automated Zoom Lenses, Ph.D. thesis, Carnegie Mellon University, 1994.

[4] B.J. Tordoff, D.W. Murray, Violating rotating camera geometry: the effect of radial distortion on self-calibration, in: Proceedings of ICPR, 2000.

[5] L. de Agapito, R. Hartley, E. Hayman, Linear self calibration of a rotating and zooming camera, in: Proceedings of the IEEE International Conference, CVPR99, 1999, pp. 15–21.

[6] J. Davis, X. Chen, Calibrating pan–tilt cameras in wide-area surveillance networks, in: Proceedings of ICCV 2003, vol. 1, 2003, pp. 144–150.

[7] Wan, D., Zhaou, J.: Stereo Vision Using Two PTZ Cameras. Computer Vision and Image Understanding 112(2), 184–194 (2008).

[8] Kumar, C. Micheloni, and C. Piciarelli, "Stereo localization using dual ptz cameras," in Proc. Int. Conf. Computer Analysis of Images and Patterns, Munster, GE, Sept. 2-4, 2009, vol. 5702, pp.1061-1069.

[9] C.Harris and M.J. Stephens. A combined corner and edge detector. In Alvey Vision Conference, pages 147-152, 1988.

[10] D. G. Lowe, Distinctive image features from scale-invariant feature points, International Journal of Computer Vision, 2004.

[11] William H. Press, Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling.Numerical Recipes in C: The Art of Scientific Computing. Cambridge UniversityPress, 1988.

[12] S.N. Sinha, M.Pollefeys. Towards calibrating a pan–tilt–zoom camera network, in: OMNIVIS 2004, 5th Workshop on Omnidirectional Vision, Camera Networks and Non-classical cameras (in conjunction with ECCV 2004), Prague, Czech Republic.

[13] Hartley, R. I.: 'Self-calibration of Stationary Cameras', International Journal of Computer Vision, 1997, 22 (1) pp. 5-23.

[14] J. Davis and X. Chen, Calibrating pan-tilt cameras in wide-area surveillance networks. In Proc. of ICCV 2003, Vol 1, page 144-150, October 2003.

[15] R. I. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, ISBN: 0521540518, second edition, 2004.

[16] R. Tsai. An efficient and accurate camera calibration technique for 3d machine vision. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 1986.

[17] Z. Zhang. A flexible new technique for camera calibration In Proceedings of 7th International Conference on Computer Vision, 1999.

[18] D. Chen and G. Zhang. A new sub-pixel detector for x-corners in camera calibration targets. WSCG Short Papers, pages 97-100, 2005.

[19] Loop, C. and Zhang, Z. (1999). Computing rectifying homographies for stereo vision.In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'99), pp. 125–131, Fort Collins.

[20] Faugeras, O. and Luong, Q.-T. (2001). The Geometry of Multiple Images. MIT Press, Cambridge, MA.